MATH5835M Practical

Jochen Voss

2023-11-24

* The MATH5835M module is assessed by an examination (80%) and a practical (20%). This is the practical, worth 20% of your final module mark.
* You must hand in your solution via Gradescope by **Thursday, 14th December 2023, 2pm**.
* Reports must be typeset (not handwritten) and should be no more than 6 pages in length.
* Within reason you may talk to your friends about this piece of work, but you should not send R code (or output) to each other. Your report must be your own work.

## Introduction

The distribution of wind speeds in a city follows a mixture of log-normal distributions: If we denote the wind speed (in miles per hour) by $X$, we have where $K$ is random with and the means and standard deviations are given by The density of the log-normal distribution with parameters $μ\in R$ and $σ>0$ is

$$φ\left(x;μ,σ\right)=\frac{1}{x\sqrt{2πσ^{2}}}exp\left(−\frac{\left(log\left(x\right)−μ\right)^{2}}{2σ^{2}}\right)$$

for all $x\geq 0$. This function is available in R as dlnorm(). You can generate samples from the wind speed distribution as follows:

wind\_speed <- function(n, w, mu, sigma) {
 k <- length(w)
 i <- sample(k, n, replace = TRUE, prob = w)
 rlnorm(n, mu[i], sigma[i])
}

w <- c(0.1, 0.4, 0.5)
mu <- c(2.1, 1.6, -0.5)
sigma <- c(0.55, 0.7, 0.6)
wind\_speed(5, w, mu, sigma)

## [1] 0.5922097 9.8571426 10.6458888 1.0282306 0.2759617

## Task 1

We are interested in the probability that the wind speed is larger than 40 miles per hour.

* Use basic Monte Carlo estimation to estimate the probability that the wind speed is larger than 40 miles per hour.
* Determine the root mean squared error of your estimate.

## Task 2

Our aim here is to find a more efficient estimator for the probability from task 1.

* Use importance sampling to estimate the probability that the wind speed is larger than 40 miles per hour.
* Compare the root mean squared error of your estimate with the one from Task 1.

## Task 3

Now, for simplicity, assume that the wind speed follows a single log-normal distribution with parameters $μ$ and $σ$: Furthermore, we assume that $μ=0$ is known and $σ$ is unknown, random and exponentially distributed: $σ∼Exp\left(1\right)$.

On five different days we have observed the wind speeds $x\_{1}=0.50$, $x\_{2}=1.67$, $x\_{3}=2.22$, $x\_{4}=0.22$ and $x\_{5}=4.36$. Using these data we want to learn about the unknown value $σ$. From Bayes rule we know that the conditional distribution of $σ$, given the data $\left(x\_{1},…,x\_{5}\right)$ has density where $Z$ is the normalising constant, $π\left(s\right)$ is the density of the $Exp\left(1\right)$ distribution, and $φ\left(x\_{i};0,s\right)$ is the density of the log-normal distribution with parameters $μ=0$ and $σ=s$. In Bayesian statistics, $π$ is called the *prior density* and $f$ is called the *posterior density*.

We want to sample from the posterior density $f$ using envelope rejection sampling.

* Produce a plot of $\tilde{f}\left(s\right)$ as a function of $s$.
* Choose an appropriate density $g\left(s\right)$ for the proposals for the rejection sampling method. Explain why your choice is appropriate.
* Show a plot of $\tilde{f}\left(s\right)/g\left(s\right)$, and from this determine an appropriate value for the constant $c$ in the envelope rejection sampling algorithm.
* Implement the envelope rejection sampling algorithm and generate $N=10,000$ samples from the posterior density $f$. Use these samples to plot a histogram.
* Compare the posterior distribution to the prior distribution, and comment on the differences you see.